Saddlepoint tests for quantile regression

Journal Item

The Canadian Journal of Statistics
Vol. xx, No. yy, 20??, Pages 1–29
La revue canadienne de statistique

Saddlepoint tests for quantile regression

Elvezio Ronchetti[1] and Radka Sabolová[2*]

[1] *Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Switzerland*
[2] *Department of Mathematics and Statistics, The Open University, Milton Keynes, UK*

Key words and phrases: Saddlepoint techniques; quantile regression; M-estimator; robust tests.

MSC 2010: Primary 62J05; secondary 62F05

Abstract: Quantile regression is a flexible and powerful technique which allows to model the quantiles of the conditional distribution of a response variable given a set of covariates. Regression quantile estimators can be viewed as $M-$estimators and standard asymptotic inference is readily available based on likelihood-ratio, Wald, and score-type test statistics. However, these statistics require the estimation of the sparsity function $s(\alpha) = [g(G^{-1}(\alpha))]^{-1}$, where G and g are the cumulative distribution function and the density of the regression errors respectively, and this can lead to nonparametric density estimation. Moreover, the asymptotic χ^2 distribution for these statistics can provide an inaccurate approximation of tail probabilities and this can lead to inaccurate p-values, especially for moderate sample sizes. Alternative methods which do not require the estimation of the sparsity function, include rank techniques and resampling methods to obtain confidence intervals, which can be inverted to test hypotheses. These are typically more accurate than the standard $M-$ tests.

In this paper we show how accurate tests can be obtained by using a nonparametric saddlepoint test statistic. The proposed statistic is asymptotically χ^2 distributed, does not require the specification of the error distribution, and does not require the estimation of the sparsity function. The validity of the method is demonstrated through a simulation study, which shows both the robustness and the accuracy of the new test compared to the best available alternatives.

Résumé: La régression par quantiles est une technique souple et puissante pour modéliser les quantiles de la distribution conditionnelle d'une variable en fonction d'un ensemble de covariées. Les estimateurs de la régression par quantiles peuvent être écrits comme des $M-$estimateurs et l'inférence asymptotique standard est disponible. Elle est basée sur des statistiques de test du type du rapport de vraisemblance, Wald et multiplicateur de Lagrange. Toutefois le calcul de ces statistiques nécessite de l'estimation de la fonction de "sparsity" $s(\alpha) = [g(G^{-1}(\alpha))]^{-1}$, où G et g sont la fonction de répartition et la fonction de densité des erreurs respectivement et ceci demande de l'estimation nonparamétrique. En plus la distribution asymptotique χ^2 de ces statistiques peut donner lieu à des approximations peu précises des probabilités dans les queues de la distribution et ceci peut produire des $p-$valeurs imprécises surtout dans le cas d'échantillons de taille moyenne.

Des méthodes alternatives qui ne nécessitent pas de l'estimation de la fonction de "sparsity" sont disponibles, par exemple les méthodes de rang et celles de reéchantillonage pour obtenir des intervalles de confiance qui peuvent être inversés afin de construire des tests. Ces méthodes sont typiquement plus précises que les $M-$tests standard.

Dans cet article on présente une statistique nonparamétrique de pointe de selle qui permet de construire des tests très précis. Elle est asymptotiquement distribuée selon une loi du χ^2 et elle ne nécessite ni de la spécification de la distribution des erreurs ni de l'estimation de la fonction de "sparsity". La performance de la méthode est démontrée par une étude de simulation, qui montre la robustesse et la précision du nouveau test en comparaison avec les meilleures alternatives disponibles dans la littérature.

1. INTRODUCTION

Introduced in a seminal paper by Koenker & Bassett (1978), quantile regression has become a standard tool in statistical methodology and practice. Instead of modelling the conditional expectation of the response given the covariates, it models the $\alpha-$quantiles of the conditional distribution and provides a richer information on the underlying relationship between the response and the covariates. From the original formulation for the standard regression model, many extensions have been provided, including generalized linear models, survival data, autoregressive models, penalized methods, and nonparametric regression. Moreover, many applications in various fields ranging from economics and finance to biology and ecology have been developed. An excellent overview on theoretical, computational, and applied aspects is given in the book Koenker (2005).

Let $Y_1, \ldots, Y_n$ be observations following the regression model

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + u_i, \qquad i = 1, \ldots, n$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \in R^p$, $x_{i1} \equiv 1$, $\boldsymbol{\beta} \in R^p$, and $u_i \sim G$ with density g. Notice that our test will not require to specify G. Later we will denote by $\mathbf{X}$ the design matrix with i-th row $\mathbf{x}_i$. Although a more general model with non-iid errors would be more useful in the context of regression quantiles, we derive our results in the simpler model given above, but we provide some numerical results for a location-scale model in Table 10, where the standard deviation of the errors depends linearly on the x's.

The regression quantile estimator $\hat{\boldsymbol{\beta}}_\alpha$ for $\boldsymbol{\beta}$ is the solution of the minimization problem

$$\hat{\boldsymbol{\beta}}_\alpha = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\alpha(Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \tag{1}$$

where

$$\rho_\alpha(u) = |u| \left\{ (1-\alpha)\, \mathrm{I}\,[u < 0] + \alpha\, \mathrm{I}\,[u > 0] \right\}.$$

It is an M-estimator defined by (9) with score function

$$\psi(y; \boldsymbol{\beta}) = \psi_\alpha(y - \boldsymbol{x}^\top \boldsymbol{\beta}), \tag{2}$$

where

$$\psi_\alpha(u) = \alpha I[u > 0] - (1-\alpha) I[u < 0] = \alpha - I[u < 0]. \tag{3}$$

The estimator $\hat{\boldsymbol{\beta}}_\alpha$ is consistent for $\boldsymbol{\beta}_\alpha = (\beta_1 + G^{-1}(\alpha), \beta_2, \ldots, \beta_p)$. Since its influence function (Hampel (1974), Hampel et al. (1986)), which is proportional to the score function (2), is bounded, the regression quantile estimator is robust against moderate deviations from the underlying error distribution when the x's are not too discordant; see He et al. (1990).

From the inferential point of view, an exact formula for the joint density of the regression quantile estimator is available; see Koenker (2005), Theorem 3.1, p. 70, Jurečková (2010), Portnoy (2012). However, from an operational point of view it requires the computation of $\binom{n}{p}$ terms and this becomes unfeasible in many applications. Accurate finite sample and saddlepoint approximations to the density of regression quantiles are available; see Spady (1991), De Jongh et al. (2004). However, the joint density would have to be marginalized to make inference on

* *Author to whom correspondence may be addressed.*
E-mail: radka.sabolova@open.ac.uk

single components, and this would be numerically very challenging. Therefore, inference on the parameters is typically carried out through the approximation provided by the asymptotic normal distribution of the estimator. Notice that the asymptotic variance (8) depends on the so-called sparsity function $s(\alpha) = [g(G^{-1}(\alpha))]^{-1}$, whose estimation is challenging and can lead to nonparametric density estimation. A variety of methods to construct confidence intervals not requiring density estimation are available. They include the inversion of rank-tests (Gutenbrunner & Jurečková (1992) and Gutenbrunner et al. (1993)), a range of resampling methods (Parzen, Wei, & Ying (1994)), Markov Chain Marginal Bootstrap (Hu & He (2012)), jackknife (Portnoy (2014)), and a "direct" method by Zhou & Portnoy (1996).

$M-$tests, i.e tests based on $M-$estimators can also be used. They are the natural counterpart of standard Wald-, score-, and likelihood-ratio tests, where - loglikelihood is replaced by the objective function $\rho_\alpha(\cdot)$ in (1), and the score function by $\psi_\alpha(\cdot)$ in (3). Their robustness and asymptotic properties have been studied in Heritier & Ronchetti (1994). It turns out that a bounded score function guarantees robustness of validity and robustness of efficiency for these tests. They all require the estimation of the sparsity function. Finally, notice that in this case score tests are a class of generalized rank tests. Therefore, rank-based inference can be carried out, but in practice one still has to rely on the asymptotic distribution of the test statistic.

$P-$values based on the asymptotic distribution of the regression quantiles estimator or of the $M-$test statistics discussed above can be misleading when the sample size is moderate and/or when small tail probabilities are required.

In this paper we focus on hypotheses testing and develop a so-called saddlepoint test, which exhibits several desirable properties especially in small samples. The test statistic is asymptotically χ^2 distributed under the null hypothesis and is therefore first-order equivalent to the standard $M-$tests. However, it exhibits a better finite sample behavior than the latter by combining excellent accuracy even in small samples and robustness. The test statistic is given by an explicit formula, is nonparametric, and it does not require the estimation of the sparsity function. It is derived from the results in Robinson, Ronchetti, & Young (2003) for $M-$estimators, which were obtained using saddlepoint techniques (Daniels (1954)) and can be viewed as an empirical likelihood procedure based on tilted exponential weights; cf. the discussion in Ma & Ronchetti (2011), p. 148. The corresponding weights are different from those obtained by the standard empirical likelihood approach by Owen (1988), Owen (2001).

The paper is organized as follows. In Section 2. we consider the case of a simple hypothesis and we derive the saddlepoint test statistic in the parametric case. This is useful to understand the construction of the new statistic. Then, we obtain the test statistic in its nonparametric version, i.e. when we do not specify the errors distribution. Section 3. is devoted to the composite hypothesis case. Compared to the simple hypothesis case where the formula is explicit, here we need an additional numerical minimization over the nuisance parameters. The simulation study of Section 4. shows the excellent accuracy and robustness properties of the nonparametric saddlepoint test statistic in finite samples. Comparisons with Wald, likelihood-ratio type, rank tests, resampling techniques, and a "direct" method are provided. Overall the χ^2_p quantiles are very close to the quantiles of the distribution of the nonparametric saddlepoint test statistic. Even in the extreme situation of 21 observations, 6 parameters, and under a spectrum of distributions for the errors, the accuracy of the new test is still good. While the standard $M-$ tests break down, the saddlepoint test has good accuracy comparable to the best resampling or rank methods. Finally, in Section 5. we provide some conclusions and discuss possible extensions of this work. In the Appendix, we summarize for completeness the definition of the saddlepoint test statistic for $M-$estimators and its properties as developed in Robinson, Ronchetti, & Young (2003) and we give the assumptions and the proofs of the Propositions.

2. SIMPLE HYPOTHESIS

Consider the simple hypothesis $H_0 : \boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_{\alpha 0}$. Although we will mostly use only its nonparametric version, it is useful to provide first the derivation of the test statistic in the parametric setup.

2.1. Parametric case

To derive the test statistic in this case, we assume for convenience that $(Y_i, \mathbf{x}_i)$ are independent identically distributed with density $g(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})k(\mathbf{x}_i)$, where $k(\cdot)$ is the density of $\mathbf{x}_i$. We will not have to specify the latter, because the final test statistic will be an expectation with respect to $k(\cdot)$ and it will simply be replaced by the average over the $\mathbf{x}_i$'s.

We now proceed to compute the test statistic $2nh(\hat{\boldsymbol{\beta}}_\alpha) = 2n \sup_\lambda\{-K_\psi(\boldsymbol{\lambda}; \hat{\boldsymbol{\beta}}_\alpha)\}$ (see Appendix), where $K_\psi(\boldsymbol{\lambda}; \hat{\boldsymbol{\beta}}_\alpha)$ is the cumulant generating function of the score function $\psi(Y_i; \boldsymbol{\beta})$ defined by (2) and (3) corresponding to the regression quantile and given by

$$
\begin{aligned}
K_\psi(\boldsymbol{\lambda}; \boldsymbol{\beta}_\alpha) &= \log \mathsf{E} e^{\boldsymbol{\lambda}\psi(Y_i;\boldsymbol{\beta}_\alpha)\mathbf{x}_i} \\
&= \log \mathsf{E} e^{\boldsymbol{\lambda}^\top \mathbf{x}_i\left(\alpha - \mathrm{I}\,[Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_\alpha < 0]\right)} \\
&= \log \int\int e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i\,\mathrm{I}\,[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_\alpha < 0]} g(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})k(\mathbf{x}_i)\,\mathrm{d}y_i\,\mathrm{d}\mathbf{x}_i \\
&= \log \int \left[\int_{-\infty}^{\mathbf{x}_i^\top \boldsymbol{\beta}_\alpha} e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i - \boldsymbol{\lambda}^\top \mathbf{x}_i} g(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})k(\mathbf{x}_i)\,\mathrm{d}y_i \right. \\
&\quad \left. + \int_{\mathbf{x}_i^\top \boldsymbol{\beta}_\alpha}^{\infty} e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} g(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})k(\mathbf{x}_i)\,\mathrm{d}y_i \right] \mathrm{d}\mathbf{x}_i \\
&= \log \int \left\{ e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right) + 1 - G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right]\right\} \mathrm{d}\mathbf{x}_i.
\end{aligned}
$$

In order to compute the saddlepoint, we need the derivative of $K_\psi(\boldsymbol{\lambda}; \boldsymbol{\beta}_\alpha)$ with respect to $\boldsymbol{\lambda}$:

$$
\begin{aligned}
&\frac{\partial K_\psi(\boldsymbol{\lambda}; \boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\lambda}} \\
&= \frac{\partial}{\partial \boldsymbol{\lambda}} \log \int \left\{ e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right) + 1 - G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right]\right\} \mathrm{d}\mathbf{x}_i \\
&= \frac{\int \left\{ \frac{\partial}{\partial \boldsymbol{\lambda}} e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right) + 1 - G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right]\right\} \mathrm{d}\mathbf{x}_i}{\int \left\{ e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})) + 1 - G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right]\right\} \mathrm{d}\mathbf{x}_i} \\
&= \left\{ \int \left\{ \alpha \mathbf{x}_i e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right) + 1 - G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right] + \right.\right. \\
&\quad \left.\left. e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[-\mathbf{x}_i e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right]\right\} \mathrm{d}\mathbf{x}_i \right\} \times \\
&\quad \left\{ \int \left\{ e^{\alpha\boldsymbol{\lambda}^\top \mathbf{x}_i} k(\mathbf{x}_i) \left[e^{-\boldsymbol{\lambda}^\top \mathbf{x}_i} G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right) + 1 - G\left(\mathbf{x}_i^\top(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})\right)\right]\right\} \mathrm{d}\mathbf{x}_i \right\}^{-1}.
\end{aligned}
$$

By solving the equation

$$\frac{\partial K_\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\lambda}} = 0$$

we obtain that $\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha)^\top \mathbf{x}_i$ must satisfy

$$\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha)^\top \mathbf{x}_i = -\log\left\{\frac{\alpha}{1-\alpha}\frac{1-G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})\right)}{G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})\right)}\right\}, \qquad i = 1, \ldots, n.$$

Therefore, under the null hypothesis $\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_{\alpha 0}$, we get

$$\begin{aligned} h(\hat{\boldsymbol{\beta}}_\alpha) &= -K_\psi(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha); \hat{\boldsymbol{\beta}}_\alpha) \\ &= -\log \int \left(\frac{G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right)}{\alpha}\right)^\alpha \left(\frac{1-G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right)}{1-\alpha}\right)^{1-\alpha} k(\mathbf{x}_i)\,\mathrm{d}\mathbf{x}_i \end{aligned}$$

and

$$2nh(\hat{\boldsymbol{\beta}}_\alpha) = -2n \log \mathsf{E}_{\mathbf{x}}\left[\left(\frac{G\left(\mathbf{x}^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right)}{\alpha}\right)^\alpha \left(\frac{1-G\left(\mathbf{x}^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right)}{1-\alpha}\right)^{1-\alpha}\right],$$

where $\boldsymbol{\beta}_0$ is a regression parameter corresponding to $\boldsymbol{\beta}_{\alpha 0}$, i.e. $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_{\alpha 0} - (G^{-1}(\alpha), 0, \ldots, 0)^\top$.

We now replace the expectation over $\mathbf{x}$ by the average over the observed $\mathbf{x}_i$'s and the next Proposition shows that the resulting test statistic is asymptotically χ^2_p under the null hypothesis.

Proposition 2.1 *Under the Assumptions given in the Appendix,*

$$\begin{aligned} &2nh(\hat{\boldsymbol{\beta}}_\alpha) = \\ &-2n \log \frac{1}{n}\sum_{i=1}^{n} \left(\frac{G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right)}{\alpha}\right)^\alpha \left(\frac{1-G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right)}{1-\alpha}\right)^{1-\alpha} \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_p, \end{aligned} \tag{4}$$

under the null hypothesis.

The proof is given the Appendix.

The test statistic for the special case of simple quantiles can be easily obtained from (4) by setting $p = 1, \mathbf{x} \equiv 1, \hat{\beta}_\alpha = F_n^{-1}(\alpha)$, the empirical quantile, and $G(\hat{\beta}_\alpha - \beta_0) = F_0(\hat{\beta}_\alpha)$, the cumulative distribution of the observations under the null hypothesis. In this case the test statistic is simply

$$2nh(\hat{\beta}_\alpha) = -2n \log\left\{\left(\frac{F_0(\hat{\beta}_\alpha)}{\alpha}\right)^\alpha \left(\frac{1-F_0(\hat{\beta}_\alpha)}{1-\alpha}\right)^{1-\alpha}\right\}. \tag{5}$$

2.2. Nonparametric case

If we do not want to specify the distribution G of the errors, we can derive a nonparametric version of the saddlepoint test statistic. Its derivation is provided in the Appendix and is given by

$$2n\hat{h}(\hat{\boldsymbol{\beta}}_\alpha) = -2\sum_{j=1}^{n} \log\left\{ \left(\frac{\sum_{i=1}^n w_{ij} I_i}{\alpha}\right)^{\alpha} \left(\frac{\sum_{i=1}^n w_{ij}(1-I_i)}{1-\alpha}\right)^{1-\alpha} \right\}, \tag{6}$$

where

$$\begin{aligned}
w_{ij} &= \frac{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}}{\sum_{k=1}^n \left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{I_{kj}(\boldsymbol{\beta}_{\alpha 0})}}, \\
r_i &= y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_\alpha, \\
I_i &= \mathrm{I}[r_i < 0], \\
I_{ij}(\boldsymbol{\beta}) &= \mathrm{I}[r_i + \mathbf{x}_j^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}) < 0] = \mathrm{I}[y_i - \mathbf{x}_j^\top \boldsymbol{\beta} < 0], \\
F_n^j(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^n I_{ij}(\boldsymbol{\beta}), \qquad i, j = 1, \ldots, n.
\end{aligned}$$

The nonparametric saddlepoint test statistic for the special case of simple quantiles ($p = 1, x_i \equiv 1$) is given by

$$2n\hat{h}(\hat{\boldsymbol{\beta}}_\alpha) = -2n \log\left\{ \left(\frac{\sum_{i=1}^n w_i I_i}{\alpha}\right)^{\alpha} \left(\frac{\sum_{i=1}^n w_i(1-I_i)}{1-\alpha}\right)^{1-\alpha} \right\}, \tag{7}$$

where

$$\begin{aligned}
w_i &= \frac{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n(\beta_{\alpha 0})}{F_n(\beta_{\alpha 0})}\right)^{\tilde{I}_i(\beta_{\alpha 0})}}{\sum_{k=1}^n \left(\frac{\alpha}{1-\alpha}\frac{1-F_n(\beta_{\alpha 0})}{F_n(\beta_{\alpha 0})}\right)^{\tilde{I}_k(\beta_{\alpha 0})}}, \\
r_i &= y_i - \hat{\beta}_\alpha, \\
I_i &= \mathrm{I}[r_i < 0] = \mathrm{I}[y_i < \hat{\beta}_\alpha], \\
\tilde{I}_i(\beta) &= \mathrm{I}[y_i < \beta], \\
F_n(\beta) &= \frac{1}{n}\sum_{i=1}^n \tilde{I}_i(\beta), \\
\hat{\beta}_\alpha &= F_n^{-1}(\alpha).
\end{aligned}$$

We prove the asymptotic distribution of the test statistic in the case of simple quantiles. The proof shows the structure of the test statistic (7), which is based on weighted sums of the $I_i's$ and

$(1 - I_i)'s$, where the weights are given by

$$w_i = \begin{cases} \frac{1}{n}\left(\frac{\alpha}{F_n(\beta_{\alpha 0})}\right) & : y_i < \beta_{\alpha 0} \\ \frac{1}{n}\left(\frac{1-\alpha}{1-F_n(\beta_{\alpha 0})}\right) & : y_i > \beta_{\alpha 0}, \end{cases}$$

which are close to $\frac{1}{n}$ when n increases. The proof for regression quantiles is similar with a more complicated notation.

Proposition 2.2 *Under the Assumptions given in the Appendix, the test statistic $2nh(\hat{\boldsymbol{\beta}}_\alpha)$ defined by (7) converges in distribution to a χ^2_1 as $n \to \infty$ under the null hypothesis.*

The proof is given the Appendix.

3. COMPOSITE HYPOTHESIS

Suppose we now want to perform a test only for the first subvector of the regression quantile, $H_0 : \boldsymbol{\beta}_{\alpha 1} = \boldsymbol{\beta}_{\alpha 10} \in \mathbb{R}^{p_1}$, where $\boldsymbol{\beta}_\alpha = (\boldsymbol{\beta}_{\alpha 1}^\top, \boldsymbol{\beta}_{\alpha 2}^\top)^\top$ and $\hat{\boldsymbol{\beta}}_\alpha = (\hat{\boldsymbol{\beta}}_{\alpha 1}^\top, \hat{\boldsymbol{\beta}}_{\alpha 2}^\top)^\top$. We consider this hypothesis for simplicity of notation, but more general hypotheses such as those on functions of the parameters can be treated, as presented in Robinson, Ronchetti, & Young (2003).

Denote by $\mathbf{x}_{i1} \in \mathbb{R}^{p_1}$ the subvector of $\mathbf{x}_i$ consisting of the first p_1 components of $\mathbf{x}_i$. Here we derive directly the nonparametric test. Let

$$\begin{aligned} \boldsymbol{\beta}^*_{\alpha 2} &= \arg\min_{\boldsymbol{\beta}_{\alpha 2}} \left\{ -\log\left(\frac{1}{n}\sum_{i=1}^n \exp\left[\boldsymbol{\lambda}^\top \psi(y_i; (\boldsymbol{\beta}_{\alpha 10}, \boldsymbol{\beta}_{\alpha 2}))\right]\right)\right\} \\ &= \arg\min_{\boldsymbol{\beta}_{\alpha 2}} \left\{ -\log\left(\frac{1}{n}\sum_{i=1}^n \exp\left[\boldsymbol{\lambda}^\top (\alpha - I[y_i - x_{i1}^\top \boldsymbol{\beta}_{\alpha 10} - x_{i2}^\top \boldsymbol{\beta}_{\alpha 2} < 0])\mathbf{x}_i\right)\right]\right\} \end{aligned}$$

and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_{\alpha 10}, \boldsymbol{\beta}_{\alpha 2})$. Then, by following the same development as in the simple hypothesis case, we can see that

$$\mathbf{x}_j^\top \boldsymbol{\mu} = \log \frac{1-\alpha}{\alpha} \frac{F_n^j(\boldsymbol{\beta}^*)}{1 - F_n^j(\boldsymbol{\beta}^*)}$$

solves the equation

$$\sum_{i=1}^n \sum_{j=1}^n (\alpha - \mathrm{I}[r_i + \mathbf{x}_j^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^*)])\mathbf{x}_j e^{(\alpha - \mathrm{I}[r_i + \mathbf{x}_j^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^*)])\mathbf{x}_j^\top \boldsymbol{\mu}} = 0,$$

and the weights are given by

$$w_{ij} = \frac{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}^*)}{F_n^j(\boldsymbol{\beta}^*)}\right)^{I_{ij}(\boldsymbol{\beta}^*)}}{\sum_{k=1}^n \left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}^*)}{F_n^j(\boldsymbol{\beta}^*)}\right)^{I_{kj}(\boldsymbol{\beta}^*)}}.$$

Hence

$$\mathbf{x}_j^\top \boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_{\alpha 1}, \boldsymbol{\beta}_{\alpha 2}) = \log\left\{\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}\right\},$$

where

$$I_i(\boldsymbol{\beta}_{\alpha 2}) = I[y_i - x_{i1}\hat{\boldsymbol{\beta}}_{\alpha 1} - x_{i2}\boldsymbol{\beta}_{\alpha 2} < 0].$$

Finally, we obtain

$$K_\psi^j((\hat{\boldsymbol{\beta}}_{\alpha 1}, \boldsymbol{\beta}_{\alpha 2}), \boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_{\alpha 1}, \boldsymbol{\beta}_{\alpha 2})) = \log\left\{\left(\frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha}\right)^\alpha \left(\frac{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1-\alpha}\right)^{1-\alpha}\right\},$$

$$h(\hat{\boldsymbol{\beta}}_{\alpha 1}) = \inf_{\boldsymbol{\beta}_{\alpha 2}} -\frac{1}{n}\sum_{j=1}^n \log\left\{\left(\frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha}\right)^\alpha \left(\frac{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1-\alpha}\right)^{1-\alpha}\right\}$$

and under null hypothesis

$$2nh(\hat{\boldsymbol{\beta}}_{\alpha 1}) = \inf_{\boldsymbol{\beta}_{\alpha 2}} -2\sum_{j=1}^n \log\left\{\left(\frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha}\right)^\alpha \left(\frac{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1-\alpha}\right)^{1-\alpha}\right\}$$

Notice that the minimization over the nuisance parameters can be computationally challenging, especially in moderate to high dimensions.

4. SIMULATION STUDY

In order to demonstrate the accuracy and the robustness of the saddlepoint tests for regression quantiles, we performed a simulation study in two setups and we compared the new test with a variety of other available tests.

In the first set of simulations, we considered a fixed balanced design, i.e. we set the i-th row of the matrix $\boldsymbol{X}$ to be $(1, \frac{i-1}{n})$. To study the behavior of the different tests across a spectrum of different distributions, we simulated the errors from a normal, a contaminated normal (obtained as a mixture of a standard normal with another normal with larger variance), a Laplace, and a logistic distribution. The true parameter value for $\boldsymbol{\beta}$ was $(3, 2)^\top$, $p = 2$, the sample sizes for the parametric case $n = 5, 10, 20, 50, 100, 300, 1000, 10000$ and for the nonparametric case $n = 21, 51, 101$. We tested the null hypothesis $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 2)^\top$. We performed 50000 simulations and we considered the quantile $\alpha = .25$ both for the parametric and the nonparametric case. We computed the saddlepoint test defined by (4) in the parametric case and by (6) in the nonparametric case and we compared them with the Wald test. In the parametric case, the tests are calibrated at the normal model, i.e. the asymptotic variance for the Wald test is computed at the normal model and the parametric saddlepoint test is computed by (4), where $G(\cdot)$ is the normal distribution. In the nonparametric case, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_\alpha$ used in the Wald test was estimated using the formula from Koenker & Bassett (1978) (implemented in the function `summary.rq` in `R`).

The results of the simulations are summarized in Table 1 and Figures 1, 2.

We plotted the percentage (out of 50000) of simulated test statistics (dots for Wald, stars for saddlepoint) that were smaller than the $\chi^2_{2;0.9}$, $\chi^2_{2;0.95}$ and $\chi^2_{2;0.99}$ respectively versus the logarithm of the sample size. The exact figures for the parametric case can be found in Table 1. The Monte Carlo standard error is always smaller than $.001$.

In the parametric case (Figure 1) both tests behave similarly with the saddlepoint test a bit better for small sample sizes. The figure also shows the robustness of both tests which are calibrated at the normal model, but behave reasonably well across a spectrum of long-tailed distributions (with the exception of the Laplace, last row). In the nonparametric case (Figure 2) the Wald test lacks accuracy everywhere. On the other side, the nonparametric saddlepoint test exhibits excellent accuracy and robustness across all the distributions and even down to small sample sizes.

	$n=5$			$n=10$			$n=20$			$n=50$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
N(0, 1)												
SAD	0.9524	0.9842	0.9980	0.9284	0.9700	0.9963	0.9142	0.9614	0.9943	0.9066	0.9554	0.9921
Wald	0.9187	0.9599	0.9907	0.9077	0.9533	0.9896	0.9028	0.9515	0.9887	0.9013	0.9510	0.9898
cont												
SAD	0.8331	0.8856	0.9552	0.8434	0.8963	0.9522	0.8577	0.9117	0.9652	0.8674	0.9247	0.9776
Wald	0.7872	0.8306	0.8798	0.8173	0.8671	0.9215	0.8416	0.8950	0.9496	0.8609	0.9173	0.9721
Log												
SAD	0.9225	0.9655	0.9949	0.8974	0.9490	0.9903	0.8880	0.9434	0.9873	0.8834	0.9397	0.9864
Wald	0.8730	0.9221	0.9692	0.8692	0.9233	0.9721	0.8728	0.9270	0.9771	0.8774	0.9326	0.9824
Lap												
SAD	0.8496	0.9130	0.9765	0.8292	0.8933	0.9613	0.8174	0.8862	0.9566	0.8005	0.8775	0.9566
Wald	0.7849	0.8405	0.9085	0.7940	0.8531	0.9210	0.7959	0.8605	0.9320	0.7916	0.8665	0.9447

	$n=100$			$n=300$			$n=1000$			$n=10000$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
N(0, 1)												
SAD	0.9041	0.9532	0.9915	0.9015	0.9515	0.9905	0.9002	0.9499	0.9900	0.8993	0.9493	0.9893
Wald	0.9012	0.9511	0.9903	0.9012	0.9504	0.9901	0.9003	0.9495	0.9900	0.8993	0.9494	0.9893
cont												
SAD	0.8706	0.9288	0.9807	0.8740	0.9322	0.9838	0.8728	0.9308	0.9828	0.8739	0.9313	0.9834
Wald	0.8679	0.9244	0.9786	0.8727	0.9309	0.9830	0.8721	0.9305	0.9822	0.8737	0.9314	0.9833
Log												
SAD	0.8816	0.9370	0.9868	0.8811	0.9378	0.9858	0.8815	0.9382	0.9858	0.8826	0.9397	0.9864
Wald	0.8780	0.9337	0.9847	0.8798	0.9366	0.9851	0.8814	0.9379	0.9856	0.8823	0.9397	0.9864
Lap												
SAD	0.7882	0.8682	0.9546	0.7822	0.8613	0.9521	0.7798	0.8604	0.9520	0.7792	0.8608	0.9529
Wald	0.7840	0.8629	0.9490	0.7803	0.8595	0.9501	0.7789	0.8601	0.9516	0.7791	0.8607	0.9529

TABLE 1: Parametric case: $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 2)^\top$, $\alpha = 0.25$, N = 50000; data generated from: N(0,1), $.8 * N(0,1) + .2 * N(0,9)$, logistic, and Laplace. The frequencies of accepting H_0 under the null hypothesis are reported.

In the second set of simulations, we focussed only on the nonparametric case, with $p = 6$, $n = 21, 31, 51$, and 50000 replicates. The i-th row of the design matrix $\mathbf{X}$ was set to $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{i6})$, where $x_{i1} = 1$ and $x_{ij} \sim U(0,1)$, $j = 2, 3, \ldots, 6$. The true value of the parameter $\boldsymbol{\beta}$ was set to $\boldsymbol{\beta} = (3, 1, 2, 3, 4, 5)^\top$. We test the null hypothesis $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 1, 2, 3, 4, 5)^\top$. The errors $u_i, i = 1, \ldots, n$ were generated from two distributions with the same α-quantile: normal distribution $N(0,1)$ and contaminated normal distribution $N(0,.)$ with $N(0,9)$ ($\varepsilon = 0.2$) and the simulations were carried out for different values of α: $\alpha = 0.1$, 0.15, 0.25 and 0.5.

In addition to the nonparametric saddlepoint test, we considered seven alternative tests.

(i) The nonparametric version of the Wald test (*Wald*), as in the previous simulation.

(ii) The likelihood ratio type test (*LR*) defined by the test statistic

$$LR_n = 2\frac{B}{A}\sum_{i=1}^{n}\left\{\rho_\alpha(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_{\alpha 0}) - \rho_\alpha(y_i - \mathbf{x}_i^\top\hat{\boldsymbol{\beta}}_\alpha)\right\},$$

where $A = \mathsf{E}[\psi_\alpha^2] = \alpha(1-\alpha)$, $B = \mathsf{E}[\psi'_\alpha] = g(G^{-1}(\alpha)) = s^{-1}(\alpha)$. The inverse B of the sparsity function was estimated using the following relationship between the asymptotic vari-

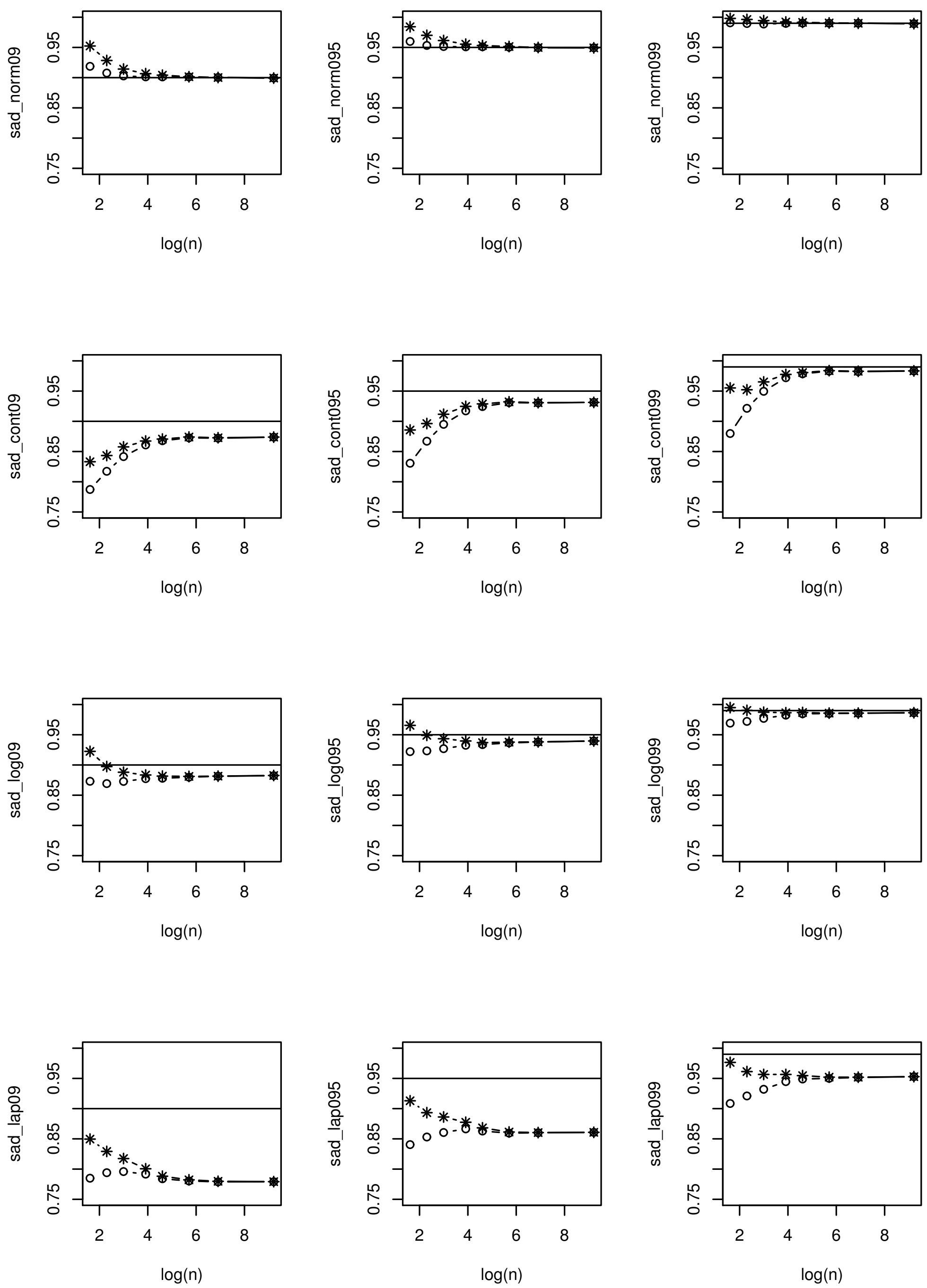

FIGURE 1: Parametric case, $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 2)^\top$, $\alpha = 0.25$, N = 50000; data generated from: N(0,1), $.8 * N(0,1) + .2 * N(0,9)$, logistic, and Laplace; dots: Wald, stars: sad. The frequencies of accepting H_0 under the null hypothesis are reported.

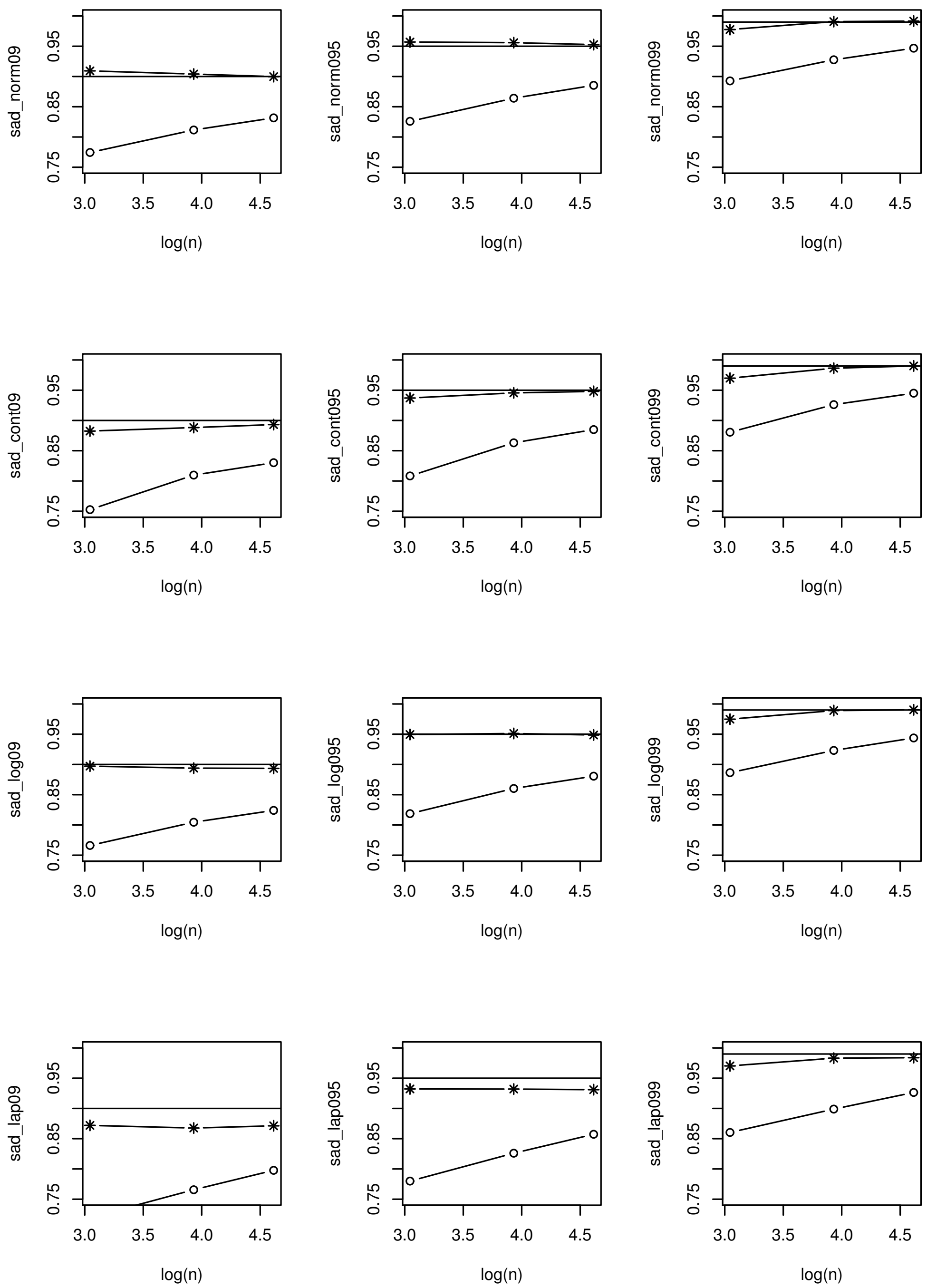

FIGURE 2: Nonarametric case, $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 2)^\top$, $\alpha = 0.25$, N = 50000; data generated from: N(0,1), $.8 * N(0,1) + .2 * N(0,9)$, logistic, and Laplace; dots: Wald, stars: sad. The frequencies of accepting H_0 under the null hypothesis are reported.

ance of $\hat{\boldsymbol{\beta}}_\alpha$ and B

$$\text{as var}(\hat{\boldsymbol{\beta}}_\alpha) = \frac{A}{B^2}(\mathbf{X}^\top\mathbf{X})^{-1}, \tag{8}$$

where asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_\alpha$ was estimated using the formula from Koenker & Bassett (1978) (implemented in the function `summary.rq` in `R`).

(iii) The test denoted by *rank-Koenker* as implemented in `R` in the package `quantreg` using the command `rq(..., se = "rank")`. This command returns confidence intervals as described in Koenker (1994), which can be inverted to get the test.

(iv) The test denoted by *rank-Sen*, a rank test defined by the test statistic

$$\sum_{i=1}^{n} x_{ij}\,\text{sign}(Y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_{\alpha 0})\left[\varphi_\alpha\left(\frac{R_i(|Y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_{\alpha 0}|)}{n+1}\right) - \bar{\varphi}_\alpha\right],$$

where

$$\varphi_\alpha(u) = \begin{cases} 0 \ : 0 < u < 1-\alpha \\ 1 : 1-\alpha < u < 1, \end{cases}$$

R_i denotes the rank, and $\bar{\varphi}_\alpha = \int_0^1 \varphi_\alpha(u)du = \alpha$.

(v) The test denoted by *asymp* based on the asymptotic distribution of regression quantiles

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}\right) = \frac{1}{\sqrt{n}g(G^{-1}(\alpha))}Q_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^\top\psi_\alpha(E_{i\alpha}) + O_p(n^{-1/4}),$$

where $E_{i\alpha} = e_i - G^{-1}(\alpha)$ and

$$\psi_\alpha(x) = \begin{cases} \alpha & : x > 0 \\ \alpha - 1 : x \le 0. \end{cases}$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p\left(0, Q^{-1}\frac{\alpha(1-\alpha)}{g^2(G^{-1}(\alpha))}\right)$$

and the test statistic is given by

$$n\frac{g^2(G^{-1}(\alpha))}{\alpha(1-\alpha)}(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0})^\top Q_n(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}) \sim \chi_p^2,$$

where $Q_n = \frac{1}{n}\mathbf{X}^\top\mathbf{X}$. The sparsity function $s(\alpha) = [g(G^{-1}(\alpha))]^{-1}$ was estimated by the kernel estimator

$$\frac{\hat{\boldsymbol{\beta}}_{n1}(\alpha + \nu_n) - \hat{\boldsymbol{\beta}}_{n1}(\alpha - \nu_n)}{2\nu_n},$$

where the value of ν_n was set to $\frac{3}{4}\alpha$ (as the optimal choice of bandwidth recommended in Dodge & Jurečková (2000) would yield negative values of $\alpha - \nu_n$ for small values of α).

Hence, the difference between this test and Wald test lies in the choice of bandwidth ν_n.

(vi) The test denoted by *direct* obtained by inverting confidence intervals constructed by using directly the empirical quantile function; see Zhou & Portnoy (1996).

(vii) The test denoted by *pivot-resam* obtained by inverting confidence intervals constructed by resampling a pivotal estimating function; see Parzen, Wei, & Ying (1994).

Under the null hypothesis, all the tests have asymptotically a χ^2_p distribution.

Tables 2, 3, 4, 5 summarize the results. Similar results are obtained for different combinations of the simulation's parameters.

norm, 21, 0.1

	0.9	0.95	0.99
sad	0.88102	0.88694	0.88870
Wald	0.10452	0.14314	0.23052
LR	0.31942	0.43646	0.63936
rank-Sen	0.76316	0.88876	0.99278
rank-Koenker	0.99464	0.99828	1
asymp	0.29802	0.31960	0.35678
direct	0.78849	0.81609	0.85392
pivot-resam	0.90026	0.96562	0.99426

norm, 21, 0.15

	0.9	0.95	0.99
sad	0.94976	0.95946	0.96288
Wald	0.28230	0.35000	0.47054
LR	0.50868	0.61882	0.78202
rank-Sen	0.93752	0.98194	0.99966
rank-Koenker	0.99172	0.99776	1
asymp	0.46778	0.49726	0.54654
direct	0.78849	0.81609	0.85392
pivot-resam	0.92590	0.97704	0.99814

norm, 31, 0.1

	0.9	0.95	0.99
sad	0.95166	0.95882	0.96166
Wald	0.19896	0.25186	0.35566
lr	0.42800	0.53526	0.71360
rank-Sen	0.64038	0.80546	0.97088
rank-Koenker	0.98356	0.99548	0.99964
asymp	0.49138	0.52192	0.57022
direct	0.73465	0.75330	0.77826
pivot-resam	0.91912	0.97194	0.99628

norm, 31, 0.15

	0.9	0.95	0.99
sad	0.97106	0.98692	0.99238
Wald	0.43368	0.49620	0.60226
LR	0.61584	0.70680	0.83724
rank-Sen	0.93480	0.97790	0.99918
rank-Koenker	0.95916	0.9863	0.9994
asymp	0.57320	0.60272	0.65146
direct	0.73465	0.75330	0.77826
pivot-resam	0.93758	0.98090	0.99834

norm, 51, 0.1

	0.9	0.95	0.99
sad	0.97328	0.98814	0.99472
Wald	0.42490	0.48230	0.58250
lr	0.60702	0.69722	0.82658
rank-Sen	0.40648	0.60694	0.88980
rank-Koenker	0.95330	0.98482	0.99902
asymp	0.63534	0.66484	0.70734
direct	0.72123	0.73592	0.75360
pivot-resam	0.91490	0.97134	0.99702

norm, 51, 0.15

	0.9	0.95	0.99
sad	0.97372	0.99034	0.99666
Wald	0.58336	0.63694	0.72626
LR	0.71636	0.79106	0.88792
rank-Sen	0.91640	0.96440	0.99532
rank-Koenker	0.93392	0.97356	0.99800
asymp	0.67636	0.70202	0.74096
direct	0.82138	0.83430	0.84858
pivot-resam	0.93578	0.98102	0.99870

TABLE 2: Nonparametric case: $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 1, 2, 3, 4, 5)^\top$, $\alpha = 0.1$ and $\alpha = 0.15$, N = 50000; data generated from N(0,1). The frequencies of accepting H_0 under the null hypothesis are reported.

In general the *Wald* and the *asymp* tests are very inaccurate even under normality and should be avoided. The *likelihood-ratio type* test is better, but still too inaccurate, except for some sample sizes and for some α. The test based on the *direct* method has a performance somewhere in the middle. The *saddlepoint* test, the especially the *rank-koenker* test and the *pivot-resam* test are the most reliable across distributions, different values of α, and even down to very small sample sizes. Notice that the *pivot-resam* is accurate, but more computational intensive than the saddlepoint test. For instance in our simulation, its computing time is 180, 100, 40 times higher than that of the saddlepoint test for $n = 21$, $n = 31$, $n = 51$, respectively.

In addition we conducted a power study to compare the proposed parametric and nonparametric saddlepoint test to other alternatives. The setup was the following.

norm, 21, 0.25	0.9	0.95	0.99
sad	0.94238	0.97100	0.98750
Wald	0.61222	0.67252	0.76412
LR	0.72826	0.80618	0.90266
rank-Sen	0.94910	0.98830	0.99994
rank-Koenker	0.97088	0.99394	0.99988
asymp	0.56892	0.59784	0.64588
direct	0.73680	0.75187	0.76335
pivot-resam	0.93466	0.98284	0.99878

norm, 21, 0.5	0.9	0.95	0.99
sad	0.67084	0.77202	0.86602
Wald	0.82674	0.87084	0.92410
LR	0.85438	0.90900	0.96492
rank-Sen	0.78614	0.89116	0.98640
rank-Koenker	0.94406	0.98442	1
asymp	0.67004	0.69622	0.73666
direct	0.73680	0.75187	0.76335
pivot-resam	0.86468	0.96256	0.99770

norm, 31, 0.25	0.9	0.95	0.99
sad	0.95240	0.97980	0.99566
Wald	0.65320	0.70376	0.78514
LR	0.75192	0.82202	0.91090
rank-Sen	0.99276	0.95248	0.99652
rank-Koenker	0.93940	0.97878	0.99884
asymp	0.62778	0.65540	0.69678
direct	0.82902	0.84279	0.85679
pivot-resam	0.94622	0.98612	0.99918

norm, 31, 0.5	0.9	0.95	0.99
sad	0.66752	0.78792	0.91822
Wald	0.79270	0.84338	0.91114
LR	0.84252	0.90114	0.96290
rank-Sen	0.78780	0.88384	0.97602
rank-Koenker	0.92888	0.97216	0.99806
asymp	0.69028	0.71806	0.75908
direct	0.82902	0.84279	0.85679
pivot-resam	0.86322	0.96344	0.99784

norm, 51, 0.25	0.9	0.95	0.99
sad	0.95102	0.98076	0.99754
Wald	0.72082	0.77684	0.85506
LR	0.80898	0.87418	0.94630
rank-Sen	0.82694	0.91946	0.98954
rank-Koenker	0.92440	0.96660	0.99610
asymp	0.69076	0.71664	0.75576
direct	0.91612	0.92679	0.93640
pivot-resam	0.94712	0.98650	0.99936

norm, 51, 0.5	0.9	0.95	0.99
sad	0.64118	0.77098	0.92192
Wald	0.79432	0.94772	0.91708
LR	0.84692	0.90626	0.96662
rank-Sen	0.79238	0.88228	0.97204
rank-Koenker	0.91386	0.96050	0.99488
asymp	0.75904	0.78602	0.82538
direct	0.96442	0.97822	0.98348
pivot-resam	0.85690	0.96140	0.99824

TABLE 3: Nonparametric case: $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 1, 2, 3, 4, 5)^\top$, $\alpha = 0.25$ and $\alpha = 0.5$, N = 50000; data generated from N(0,1). The frequencies of accepting H_0 under the null hypothesis are reported.

The true value of $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ was $(1/2, 1/2, 1/2)$ in the parametric case and $(1, 1, 1)$ in the nonparametric case and we tested the simple null hypothesis $H_0 : \boldsymbol{\beta}_\alpha = (0, 0, 0)$ and the composite hypothesis $H_0 : \boldsymbol{\beta}_{\alpha 2} = (\beta_{\alpha 2}, \beta_{\alpha 3}) = (0, 0)^\top$. As in the simulations under the null hypothesis, the errors u_i were generated from a standardized normal distribution N(0,1), a Student t_3 and a contaminated normal distribution. We considered values $\alpha = 0.2$ and $\alpha = 0.4$ sample sizes $n = 25$ and $n = 45$ for the parametric case, $n = 25$ for the nonparametric case, and the number of replications was set to 50000. We compared the performance of the parametric saddlepoint test to the Wald test calibrated at the normal distribution of the errors, whereas in the nonparametric case we compared the performance of the tests considered in the simulations under the null hypothesis.

Results are summarized in the Tables 6 to 9, where we present the proportion of rejections under the alternative at significance level 0.05. In the case of a simple hypothesis, the saddlepoint test shows good power, with the Wald test improving rapidly when moving toward the centre of the distribution. In the case of composite hypothesis, the power was smaller across all the tests.

cont, 21, 0.1	0.9	0.95	0.99
sad	0.70362	0.78770	0.86418
Wald	0.03148	0.05012	0.10260
LR	0.23376	0.33026	0.52510
rank-Sen	0.59876	0.75668	0.95966
rank-Koenker	0.99106	0.9975	1
asymp	0.34594	0.36948	0.40810
direct	0.79168	0.81515	0.84841
pivot-resam	0.89620	0.96556	0.99474

cont, 21, 0.15	0.9	0.95	0.99
sad	0.81418	0.87836	0.93316
Wald	0.18670	0.24184	0.35832
LR	0.45784	0.56820	0.74352
rank-Sen	0.88230	0.95762	0.99834
rank-Koenker	0.98984	0.99694	1
asymp	0.52920	0.55738	0.60396
direct	0.79168	0.81515	0.84841
pivot-resam	0.92642	0.97640	0.99822

cont, 31, 0.1	0.9	0.95	0.99
sad	0.73562	0.81704	0.92078
Wald	0.07012	0.09888	0.17024
LR	0.29808	0.39914	0.58788
rank-Sen	0.41010	0.58102	0.86088
rank-Koenker	0.98018	0.99480	0.99964
asymp	0.55898	0.58638	0.63058
direct	0.73312	0.74822	0.76997
pivot-resam	0.91552	0.97066	0.99594

cont, 31, 0.15	0.9	0.95	0.99
sad	0.82706	0.89412	0.96670
Wald	0.28308	0.34378	0.45642
LR	0.53440	0.63836	0.78894
rank-Sen	0.83168	0.92388	0.99198
rank-Koenker	0.95638	0.98534	0.99942
asymp	0.64688	0.67262	0.71232
direct	0.73312	0.74822	0.76997
pivot-resam	0.93694	0.98096	0.99860

cont, 51, 0.1	0.9	0.95	0.99
sad	0.71602	0.799	0.91050
Wald	0.18766	0.23534	0.33052
LR	0.43034	0.5327	0.70254
rank-Sen	0.15522	0.28156	0.59012
rank-Koenker	0.94996	0.98288	0.99854
asymp	0.72784	0.74862	0.78144
direct	0.75680	0.76914	0.78585
pivot-resam	0.91476	0.97114	0.99654

cont, 51, 0.15	0.9	0.95	0.99
sad	0.83258	0.89372	0.96012
Wald	0.44420	0.50370	0.60680
LR	0.64754	0.73154	0.85204
rank-Sen	0.72378	0.84552	0.96624
rank-Koenker	0.93408	0.97292	0.99792
asymp	0.77206	0.78944	0.81784
direct	0.83456	0.84699	0.86221
pivot-resam	0.93740	0.98080	0.99838

TABLE 4: Nonparametric case: $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 1, 2, 3, 4, 5)^\top$, $\alpha = 0.1$ and $\alpha = 0.15$, N = 50000; data generated from a contaminated distribution $.8 * N(0,1) + .2 * N(0,9)$. The frequencies of accepting H_0 under the null hypothesis are reported.

cont, 21, 0.25	0.9	0.95	0.99	cont, 21, 0.5	0.9	0.95	0.99
sad	0.86140	0.91612	0.95962	sad	0.67124	0.77102	0.86570
Wald	0.55818	0.62516	0.72748	Wald	0.83022	0.87384	0.92676
LR	0.72508	0.80318	0.90290	LR	0.86116	0.91362	0.96852
rank-Sen	0.95458	0.98918	0.99998	rank-Sen	0.78462	0.88988	0.98530
rank-Koenker	0.97114	0.99376	0.99974	rank-Koenker	0.94276	0.98452	1
asymp	0.64082	0.66730	0.70734	asymp	0.67704	0.70378	0.74338
direct	0.74345	0.75724	0.76851	direct	0.74345	0.75724	0.76851
pivot-resam	0.94012	0.98322	0.99878	pivot-resam	0.85938	0.96054	0.99696

cont, 31, 0.25	0.9	0.95	0.99	cont, 31, 0.5	0.9	0.95	0.99
sad	0.88332	0.93580	0.98084	sad	0.67596	0.79316	0.91926
Wald	0.60114	0.66170	0.75316	Wald	0.79900	0.84860	0.91242
LR	0.74636	0.81888	0.90754	LR	0.84884	0.90512	0.96458
rank-Sen	0.91840	0.96750	0.99766	rank-Sen	0.78676	0.88340	0.97644
rank-Koenker	0.93912	0.97712	0.99890	rank-Koenker	0.92742	0.97180	0.99838
asymp	0.70808	0.73022	0.76412	asymp	0.69238	0.71918	0.75932
direct	0.83149	0.84522	0.85876	direct	0.83149	0.84522	0.85876
pivot-resam	0.94810	0.98696	0.99924	pivot-resam	0.86264	0.96406	0.99778

cont, 51, 0.25	0.9	0.95	0.99	cont, 51, 0.5	0.9	0.95	0.99
sad	0.89326	0.94330	0.98456	sad	0.64768	0.77502	0.92540
Wald	0.69184	0.75050	0.83468	Wald	0.79982	0.85306	0.92022
LR	0.80168	0.86784	0.94262	LR	0.85300	0.91078	0.97010
rank-Sen	0.91170	0.96184	0.99556	rank-Sen	0.79184	0.88222	0.97108
rank-Koenker	0.91984	0.96516	0.99614	rank-Koenker	0.91332	0.96064	0.99502
asymp	0.76258	0.78204	0.81068	asymp	0.76200	0.78906	0.82560
direct	0.90574	0.91677	0.92709	direct	0.96329	0.97664	0.98189
pivot-resam	0.94464	0.98598	0.99908	pivot-resam	0.85566	0.96084	0.99790

TABLE 5: Nonparametric case: $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 1, 2, 3, 4, 5)^\top$, $\alpha = 0.25$ and $\alpha = 0.5$, N = 50000; data generated from $.8 * N(0,1) + .2 * N(0,9)$. The frequencies of accepting H_0 under the null hypothesis are reported.

$n = 25$

$\alpha = 0.2$	norm	t_3	cont	$\alpha = 0.4$	norm	t_3	cont
sad	0.91320	0.82402	0.90748	sad	0.94418	0.91766	0.94170
Wald	0.10292	0.26656	0.20402	Wald	0.75394	0.73732	0.75936

$n = 45$

$\alpha = 0.2$	norm	t_3	cont	$\alpha = 0.4$	norm	t_3	cont
sad	0.99164	0.93694	0.98166	sad	0.99794	0.99200	0.99658
Wald	0.13824	0.28180	0.21940	Wald	0.94422	0.91688	0.94048

TABLE 6: Parametric case, simple hypothesis: $H_0 : \boldsymbol{\beta}_\alpha = (0,0,0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\boldsymbol{\beta} = (1,1,1)^\top$, $N = 50000$; data generated from N(0,1), Student t_3-distribution and contaminated normal distribution. Powers are reported.

$n = 25$

$\alpha = 0.2$	norm	t_3	cont	$\alpha = 0.4$	norm	t_3	cont
sad	0.06362	0.47420	0.43626	sad	0.07422	0.51504	0.49676
Wald	0.08254	0.51522	0.48274	Wald	0.09120	0.55610	0.54142

$n = 45$

$\alpha = 0.2$	norm	t_3	cont	$\alpha = 0.4$	norm	t_3	cont
sad	0.09560	0.71156	0.71078	sad	0.11360	0.81138	0.82262
Wald	0.11064	0.72944	0.73160	Wald	0.12894	0.82760	0.83966

TABLE 7: Parametric case, composite hypothesis: $H_0 : \boldsymbol{\beta}_{\alpha 2} = (0, 0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\boldsymbol{\beta} = (1, 1, 1)^\top$, $N = 50000$; data generated from N(0,1), Student t_3-distribution and contaminated normal distribution. Powers are reported.

$n = 25$

$\alpha = 0.2$	norm	t_3	cont	$\alpha = 0.4$	norm	t_3	cont
sad	0.88102	0.54892	0.67534	sad	0.99996	0.99196	0.99594
Wald	0.87582	0.72790	0.83356	Wald	0.99852	0.98074	0.99234
LR	0.88612	0.66406	0.79524	LR	0.99982	0.98960	0.99660
rank-Sen	0.06882	0.04376	0.06104	rank-Sen	0.08920	0.07386	0.09212
rank-Koenker	0.07218	0.05154	0.06368	rank-Koenker	0.18124	0.13412	0.16560
asymp	0.67578	0.43348	0.47370	asymp	0.94112	0.77542	0.79958
direct	0.21448	0.52321	0.48311	direct	0.13628	0.33350	0.30094
pivot-resam	0.54452	0.46506	0.51512	pivot-resam	0.68614	0.61708	0.67806

TABLE 8: Nonparametric case, simple hypothesis: $H_0 : \boldsymbol{\beta}_{\alpha} = (0, 0, 0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\boldsymbol{\beta} = (1, 1, 1)^\top$, $N = 50000$; data generated from N(0,1), Student t_3 and a contaminated normal distribution. Powers are reported.

$n = 25$

$\alpha = 0.2$	norm	t_3	cont	$\alpha = 0.4$	norm	t_3	cont
sad	0.06316	0.04940	0.12554	sad	0.18574	0.08764	0.20028
Wald	0.36440	0.29050	0.39832	Wald	0.31540	0.19194	0.29482
LR	0.30802	0.18092	0.30494	LR	0.29486	0.15156	0.27478
rank-Sen	0.01092	0.00114	0.00570	rank-Sen	0.01948	0.00154	0.01184
rank-Koenker	0.12304	0.08408	0.10540	rank-Koenker	0.20250	0.15884	0.18792
asymp	0.31582	0.20110	0.22142	asymp	0.32340	0.23004	0.26686
direct	0.07175	0.11278	0.12255	direct	0.05326	0.08064	0.08240
pivot-resam	0.50364	0.45084	0.48188	pivot-resam	0.42270	0.37992	0.40788

TABLE 9: Nonparametric case, composite hypothesis: $H_0 : \boldsymbol{\beta}_{\alpha 2} = (0, 0)^\top$, $\alpha = 0.2$ and $\alpha = 0.4$, $\boldsymbol{\beta} = (1, 1, 1)^\top$, $N = 50000$; data generated from N(0,1), Student t_3 and a contaminated normal distribution. Powers are reported.

	$n=5$			$n=10$			$n=20$			$n=50$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
N(0,1)												
sad	0.81176	0.90326	0.98280	0.71936	0.81474	0.93492	0.62260	0.71882	0.86618	0.46230	0.55730	0.72084
Wald	0.71936	0.80088	0.90468	0.66294	0.74862	0.86390	0.58544	0.67284	0.80688	0.44226	0.52910	0.68070
cont												
sad	0.70910	0.78418	0.88798	0.65204	0.73856	0.85880	0.58458	0.67462	0.81152	0.44526	0.53496	0.68866
Wald	0.63618	0.70004	0.78544	0.60384	0.67820	0.78490	0.54938	0.63082	0.75460	0.42438	0.5092	0.64918
log												
sad	0.76974	0.86034	0.96254	0.68220	0.77792	0.90316	0.59590	0.69196	0.83750	0.45074	0.54036	0.69862
Wald	0.68158	0.75786	0.85868	0.62886	0.71268	0.82730	0.56126	0.64612	0.77650	0.43128	0.51256	0.65806
Lap												
sad	0.68672	0.77726	0.90854	0.61712	0.70604	0.83542	0.54044	0.62688	0.76306	0.41942	0.50016	0.63972
Wald	0.60566	0.67152	0.77138	0.56912	0.64312	0.75214	0.50998	0.58638	0.70306	0.40310	0.47720	0.60362

TABLE 10: Parametric case, location-scale model $Y_i = X_i^\top \boldsymbol{\beta} + X_i^\top \boldsymbol{\gamma} u_i, \boldsymbol{\gamma} = (1,1)$: $H_0 : \boldsymbol{\beta}_\alpha = (3 + G^{-1}_{N(0,1)}(\alpha), 2)^\top$, $\alpha = 0.25$, N = 50000; data generated from: N(0,1), $.8 * N(0,1) + .2 * N(0,9)$, logistic, and Laplace. The frequencies of accepting H_0 under the null hypothesis are reported.

Finally, Table 10 presents the frequencies of accepting H_0 under the null hypothesis in a location-scale model (non-iid case).

5. CONCLUSIONS AND OUTLOOK

We introduced a new test for quantile regression. It is derived using saddlepoint methods and it shows good accuracy in small sample sizes as well as good robustness properties. Although in theory the test can be inverted to obtain confidence intervals, this seems to be possible at present only by brute computational force. It would be interesting to find a direct way to obtain confidence intervals, but this is an open problem. Finally, the structure of the test is quite general and extensions can be worked out easily, including e.g. extreme regression quantiles; see Smith (1994), Portnoy & Jurečková (1999), Knight (2001), Chernozhukov (2005), and Jurečková (2007).

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, the Associate Editor, and two Referees for very helpful comments and references and Jana Jurečková for stimulating discussions. A grant of the Swiss National Science Foundation Sciex Program 10.172 is gratefully acknowledged. The work of the second author was also partially supported by the Czech Republic Grant 201/12/0083.

BIBLIOGRAPHY

Chernozhukov, V. (2005). Extremal Quantile Regression. *The Annals of Statistics*, 33, 806–839.

Daniels, H.E. (1954). Saddlepoint Approximations in Statistics. *Annals of Mathematical Statistics*, 25, 631-650.

De Jongh, P.J., De Wet, T., & Van Deventer, P.J.U. (2004). Saddlepoint Approximations for the Distribution of Regression Quantiles. *South African Statistical Journal*, 38, 57–76.

Dodge, Y., & Jurečková, J. (2000). *Adaptive Regression*, Springer, New York.

Gutenbrunner, C., & Jurečková, J. (1992). Regression Rank Scores and Regression Quantiles. *The Annals of Statistics*, 20, 305–330.

Gutenbrunner, C., Jurečková, J., Koenker, R., & Portnoy, S. (1993). Tests of Linear Hypotheses Based on Regression Rank Scores. *Journal of Nonparametric Statistics*, 2, 307–331.

Hampel, F.R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69, 383–393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., & Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

He, X., Jurečková, J., Koenker, R., & Portnoy, S. (1990). Tail Behavior of Regression Estimators and their Breakdown Points. *Econometrica*, 58, 1195–1214.

Heritier, S., & Ronchetti, E. (1994). Robust Bounded-Influence Tests in General Parametric Models. *Journal of the American Statistical Association*, 89, 897–904.

Hu, J., & He, X. (2012). Searching for Alternative Splicing With a Joint Model on Probe Measurability and Expression Intensities. *Journal of the American Statistical Association*, 107, 935–945.

Jurečková, J. (2007). Remark on Extreme Regression Quantile. *Sankhyā: The Indian Journal of Statistics*, 69, 87–100.

Jurečková, J. (2010). Finite-Sample Distribution of Regression Quantiles. *Statistics & Probability Letters*, 80, 1940–1946.

Knight, K. (2001). Limiting Distributions of Linear Programming Estimators. *Extremes*, 4, 87–103.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33–50.

Koenker, R. W. (1994). Confidence Intervals for Regression Quantiles. In Hušková, M. and Mandl, P. (Eds.) *Asymptotic Statistics*, Springer-Verlag, New York, 349–359.

Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.

Lô, S.N., & Ronchetti, E. (2009). Robust and Accurate Inference for Generalized Linear Models. *Journal of Multivariate Analysis*, 100, 2126–2136.

Ma, Y., & Ronchetti, E. (2011). Saddlepoint Test in Measurement Error Models. *Journal of the American Statistical Association*, 106, 147–156.

Owen, A.B. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, 75, 237–249.

Owen, A.B. (2001). *Empirical Likelihood*, Chapman and Hall/CRC.

Parzen, M.I., Wei, L.J. & Ying, Z. (1994). A Resampling Method Based on Pivotal Estimating Functions. *Biometrika*, 81, 341–350.

Portnoy, S., & Jurečková, J. (1999). On Extreme Regression Quantiles. *Extremes*, 2, 227–243.

Portnoy, S. (2012). Nearly Root-n Approximation for Regression Quantile Processes. *The Annals of Statistics*, 40, 1714–1736.

Portnoy, S. (2014). The Jackknife's Edge: Inference for Censored Regression Quantiles. *Computational Statistics and Data Analysis*, 72, 273–281.

Robinson, J., Ronchetti, E., & Young, G.A. (2003). Saddlepoint Approximations and Tests Based on Multivariate M-estimates. *The Annals of Statistics*, 31, 1154–1169.

Ronchetti, E., & Welsh, A.H. (1994). Empirical Saddlepoint Approximations for Multivariate M-estimators. *Journal of the Royal Statistical Society*, 56, 313–326.

Smith, R.L. (1994). Nonregular Regression. *Biometrika*, 81, 173–183.

Spady, R.H. (1991). Saddlepoint Approximations for Regression Models. *Biometrika*, 78, 879–889.

Zhou, K.Q., & Portnoy, S. (1996). Direct Use of Regression Quantiles to Construct Confidence Sets in Linear Models. *The Annals of Statistics*, 24, 287–306.

APPENDIX

Assumptions

A1 The distribution function G is absolutely continuous, with continuous density g differentiable at the point $G^{-1}(\alpha)$ and uniformly bounded away from 0 and ∞ at the point $G^{-1}(\alpha)$.

A2 There exist positive definite matrices $\mathbf{D}_0$ and $\mathbf{D}_1(\alpha)$ such that

(i) $\lim_{n\to\infty} \frac{1}{n}\mathbf{X}^\top\mathbf{X} = \mathbf{D}_0$,

(ii) $\lim \frac{1}{n}\sum_{i=1}^{n} g(G^{-1}(\alpha))\mathbf{x}_i\mathbf{x}_i^\top = \mathbf{D}_1(\alpha)$,

(iii) $\max_{i=1,\dots,n} ||\mathbf{x}_i||/\sqrt{n} \to 0$

A3 We assume $\sqrt{n}(a_n(\varepsilon) - \alpha) \to \infty$ and $\sqrt{n}(b_n(\varepsilon) - \alpha) \to \infty$, where

$$a_n(\varepsilon) = \frac{1}{n} \sum_{i=1}^{n} G_{ni}(\mathbf{x}_i^\top \boldsymbol{\beta}_{\alpha 0} - \varepsilon)$$

$$b_n(\varepsilon) = \frac{1}{n} \sum_{i=1}^{n} G_{ni}(\mathbf{x}_i^\top \boldsymbol{\beta}_{\alpha 0} + \varepsilon)$$

and G_{ni} denotes the conditional distribution function of Y_i, $i = 1, 2, \ldots, n$.

A4 There exists $d > 0$ such that

$$\liminf_{n\to\infty} \inf_{||u||=1} \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}[|\mathbf{x}_i^\top u| < d] = 0.$$

A5 There exists $D > 0$ such that

$$\limsup_{n\to\infty} \sup_{||u||=1} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^\top u)^2 \leq D.$$

Assumptions **A1** - **A2** are conditions for asymptotic normality of regression quantiles; see Koenker (2005), section 4.2. The assumption **A1** is a strengthened version of the condition A1 in Koenker (2005) in that we require the differentiability of the density function, which is necessary for the Taylor expansion of the test statistic. Assumptions **A3** - **A5** are conditions for the consistency of a regression quantile; see Koenker (2005), section 4.1.2.

Saddlepoint Test for M-estimators

For completeness we summarize here the definition of the saddlepoint test statistic for $M-$estimators and its properties as developed in Robinson, Ronchetti, & Young (2003).

Let $Y_1, \ldots, Y_n$ be an independent sample from a distribution F. Consider an M-estimator $\hat{\theta}$ of a parameter $\theta = \theta(F)$, defined as a solution of the equation

$$\sum_{i=1}^{n} \psi(Y_i; \theta) = 0. \tag{9}$$

Consider the composite hypothesis

$$H_0 : \theta_1 = \theta_{10} \in \mathbb{R}^{p_1}, \theta_2 \in R^{p_2}$$

where $\theta = (\theta_1^\top, \theta_2^\top), \hat{\theta} = (\hat{\theta}_1^\top, \hat{\theta}_2^\top)$.

Parametric case

Define the one-dimensional statistic

$$h(\hat{\theta}_1) = \inf_{\theta_2} \sup_{\lambda} \{-K_\psi(\lambda; (\hat{\theta}_1, \theta_2))\}, \tag{10}$$

where

$$K_\psi(\lambda; \theta) = \log \mathsf{E}_{F_0}[e^{\lambda^\top \psi(Y_i;\theta)}]$$

is the cumulant generating function of the score $\psi(Y_i;\theta)$ and the expectation is computed with respect to the distribution F_0 of the observations under the null hypothesis. Notice that the $\sup$ part in (10) can be rewritten as

$$\sup_{\lambda}\{-K_\psi(\lambda;t)\} = -K_\psi(\lambda(t);t), \tag{11}$$

where $\lambda(t)$ is the saddlepoint satisfying

$$\frac{\partial}{\partial\lambda}K_\psi(\lambda;t) = 0,$$

i.e.

$$\mathsf{E}_{F_0}[\psi(Y_i;t)e^{\lambda^\top\psi(Y_i;t)}] = 0.$$

Then under the null hypothesis,

$$2nh(\hat{\theta}_1) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p_1}$$

with relative error of order $O(n^{-1})$. This test is first-order equivalent to the three classical tests, but exhibits better second-order properties, i.e. has better small sample properties.

In the case of a simple hypothesis, the statistic simplifies to

$$h(\hat{\theta}) = -K_\psi(\lambda(\hat{\theta});\hat{\theta}).$$

Moreover, if the observations $Y_1,\ldots,Y_n$ are independent but not identically distributed (as in the regression case), the cumulant generating function becomes

$$K_\psi(\lambda;\theta) = \frac{1}{n}\sum_{i=1}^{n}K^i_\psi(\lambda;\theta),$$

where $K^i_\psi(\lambda;\theta) = \log\mathsf{E}_{F^i}[e^{\lambda^\top\psi(Y_i;\theta)}]$ and F^i is the distribution function of Y_i; see Lô & Ronchetti (2009).

Nonparametric case

When F is unspecified, an empirical version of the test may be used. Let $\hat{F}_0 = (w_1,\ldots,w_n)$ be the empirical distribution which satisfies the null hypothesis and is closest to $(1/n,\ldots,1/n)$ in the sense of the backward Kullback-Leibler divergence.

Then, the saddlepoint test statistic is given by

$$\hat{h}(\hat{\theta}_1) = \inf_{\theta_2}\sup_{\lambda}\left\{-K^w_0(\lambda;(\hat{\theta}_1,\theta_2))\right\},$$

where

$$K^w_0(\lambda;\theta) = \log\left(\sum_{i=1}^{n}w_ie^{\lambda^\top\psi(y_i;\theta)}\right),$$

and the weights w_i are computed as

$$w_i = e^{\mu(\theta^*)^\top \psi(y_i;\theta^*)} / \sum_{j=1}^{n} e^{\mu(\theta^*)^\top \psi(y_j;\theta^*)}, \tag{12}$$

where

$$\theta^* = (\theta_{10}, \theta_2^*)$$

$$\theta_2^* = \arg\min_{\theta_2}\{-\kappa(\mu(\theta_{10},\theta_2);(\theta_{10},\theta_2))\}$$

$$\mu(\theta) = \arg\max_{\mu}\{-\kappa(\mu;\theta)\},$$

$$\kappa(\mu;\theta) = \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mu(\theta)^\top \psi(y_i;\theta)}\right).$$

When $n \to \infty$, the p-value satisfies

$$P_{H_0}\{2n\hat{h}(\hat{\theta}_1) \geq 2n\hat{h}(\hat{\theta}_{1obs})\} = \{1 - Q_{p_1}(2n\hat{h}(\hat{\theta}_{1obs}))\{1 + O_P(n^{-1})\}\},$$

where Q_{p_1} denotes the cumulative distribution function of the χ^2 distribution with p_1 degrees of freedom.

Proof of Proposition 2.1

Proof.
By Taylor expansion we obtain,

$$\begin{aligned}
G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)\right) =& G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0} + (G^{-1}(\alpha),0,\ldots,0)^\top)\right) \\
=& G(\mathbf{x}_i^\top(G^{-1}(\alpha),0,\ldots,0)^\top) \\
&+ \mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}) g(\mathbf{x}_i^\top(G^{-1}(\alpha),0,\ldots,0)^\top) \\
&+ \frac{1}{2}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}))^2 g'(\mathbf{x}_i^\top(G^{-1}(\alpha),0,\ldots,0)^\top) \\
&+ O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0} - \hat{\boldsymbol{\beta}}_\alpha))^3) \\
=& \alpha + \mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}) g(G^{-1}(\alpha)) + \\
& \frac{1}{2}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}))^2 g'(\mathbf{x}_i^\top(G^{-1}(\alpha),0,\ldots,0)^\top) \\
&+ O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0} - \hat{\boldsymbol{\beta}}_\alpha))^3)
\end{aligned}$$

and by further Taylor expansion of $(1+x)^\alpha$ we get

$$
\begin{aligned}
\left(\frac{G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_0)\right)}{\alpha}\right)^\alpha =& \Bigg(1+\frac{\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha))}{\alpha} \\
&+\frac{1}{2\alpha}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0}))^2 g'(G^{-1}(\alpha)) \\
&+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3)\Bigg)^\alpha \\
=&1+\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)) \\
&+\frac{1}{2\alpha}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0}))^2 g'(G^{-1}(\alpha)) \\
&+\frac{\alpha-1}{2\alpha}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)))^2 \\
&+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\left(\frac{1-G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_0)\right)}{1-\alpha}\right)^{1-\alpha} =& \Bigg(1-\frac{\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha))}{1-\alpha} \\
&+\frac{1}{2(1-\alpha)}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0}))^2 g'(G^{-1}(\alpha)) \\
&+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3)\Bigg)^{1-\alpha} \\
=&1-\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)) \\
&-\frac{1}{2\alpha}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0}))^2 g'(G^{-1}(\alpha)) \\
&-\frac{\alpha-1}{2\alpha}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)))^2 \\
&+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3),
\end{aligned}
$$

and finally

$$
\begin{aligned}
&\left(\frac{G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_0)\right)}{\alpha}\right)^\alpha \left(\frac{1-G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_0)\right)}{1-\alpha}\right)^{1-\alpha} \\
&=1-\frac{1}{2\alpha(1-\alpha)}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)))^2+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&-2n\log\frac{1}{n}\sum_{i=1}^{n}\left(\frac{G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_0)\right)}{\alpha}\right)^{\alpha}\left(\frac{1-G\left(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_0)\right)}{1-\alpha}\right)^{1-\alpha}\\
&=-2n\log\frac{1}{n}\sum_{i=1}^{n}\Big(1-\frac{1}{2\alpha(1-\alpha)}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)))^2\\
&\quad+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3)\Big)\\
&=-2n\log\left(1-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\alpha(1-\alpha)}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g(G^{-1}(\alpha)))^2\right.\\
&\quad\left.+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha)^3))\right)\\
&=\sum_{i=1}^{n}\frac{1}{\alpha(1-\alpha)}(\mathbf{x}_i^\top(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0}))^2+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3)\\
&=n\frac{g^2(G^{-1}(\alpha))}{\alpha(1-\alpha)}(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha)^\top\left(\sum_{i=1}^{n}\frac{1}{n}\mathbf{x}_i\mathbf{x}_i^\top\right)(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha)\\
&\quad+O_P((\mathbf{x}_i^\top(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha))^3).
\end{aligned}
$$

The proof can be completed by using the consistency and asymptotic normality of the regression quantile estimator

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})\xrightarrow[n\to\infty]{\mathcal{D}}N_p\left(0,\frac{\alpha(1-\alpha)}{g(G^{-1}(\alpha))^2}\mathbf{D}_0^{-1}\right),
$$

which hold under the Assumptions of the Proposition. Thus, the test statistic $2nh(\hat{\boldsymbol{\beta}}_\alpha)$ is asymptotically χ^2_p.

■

Proof of Proposition 2.2

Proof.

Let us rewrite the test statistic (7) and let us first consider the denominator of w_i:

$$\begin{aligned}
&\sum_{k=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n(\beta_{\alpha 0})}{F_n(\beta_{\alpha 0})}\right)^{\tilde{I}_k(\beta_{\alpha 0})} \\
&\quad = \sum_{k=1}^{n}(1-\tilde{I}_k(\beta_{\alpha 0})) + \left(\frac{\alpha}{1-\alpha}\frac{1-F_n(\beta_{\alpha 0})}{F_n(\beta_{\alpha 0})}\right)\sum_{k=1}^{n}\tilde{I}_k(\beta_{\alpha 0}) \\
&\quad = n + \left\{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n(\beta_{\alpha 0})}{F_n(\beta_{\alpha 0})}\right) - 1\right\}\sum_{k=1}^{n}\tilde{I}_k(\beta_{\alpha 0}) \\
&\quad = n + \left\{\frac{\alpha - F_n(\beta_{\alpha 0})}{(1-\alpha)F_n(\beta_{\alpha 0})}\right\} nF_n(\beta_{\alpha 0}) \\
&\quad = n\,\frac{1-F_n(\beta_{\alpha 0})}{1-\alpha}.
\end{aligned}$$

Thus,

$$\begin{aligned}
w_i I_i &= \left(\frac{\alpha}{nF_n(\beta_{\alpha 0})}\right)^{\tilde{I}_i(\beta_{\alpha 0})}\left(\frac{1-\alpha}{n(1-F_n(\beta_{\alpha 0}))}\right)^{1-\tilde{I}_i(\beta_{\alpha 0})} I_i \\
&= \frac{1}{n}\left(\frac{\alpha}{F_n(\beta_{\alpha 0})}\right)^{\tilde{I}_i(\beta_{\alpha 0})}\left(\frac{1-\alpha}{1-F_n(\beta_{\alpha 0})}\right)^{1-\tilde{I}_i(\beta_{\alpha 0})} I_i
\end{aligned}$$

and

$$\sum_{i=1}^{n} w_i I_i = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\alpha}{F_n(\beta_{\alpha 0})}\right)^{\tilde{I}_i(\beta_{\alpha 0})}\left(\frac{1-\alpha}{1-F_n(\beta_{\alpha 0})}\right)^{1-\tilde{I}_i(\beta_{\alpha 0})} I_i.$$

Similarly for $\sum_{i=1}^{n} w_i(1-I_i)$. Now

$$\begin{aligned}
E\{\tilde{I}_i(\beta_{\alpha 0})\} &= E\{I[y_i < \beta_{\alpha 0}]\} = E\{I[y_i < \beta_0 + G^{-1}(\alpha)]\} \\
&= P[u_i < G^{-1}(\alpha)] = G(G^{-1}(\alpha)) = \alpha
\end{aligned}$$

and

$$\begin{aligned}
E\{I_i\} &= E\{I[y_i < \hat{\boldsymbol{\beta}}_\alpha]\} \\
&= E\{I[y_i - \beta_0 < \hat{\boldsymbol{\beta}}_\alpha - \beta_0]\} \\
&= P[u_i < \hat{\boldsymbol{\beta}}_\alpha - \beta_0] = G(\hat{\boldsymbol{\beta}}_\alpha - \beta_0).
\end{aligned}$$

By the law of large numbers $F_n(\beta_{\alpha 0}) \to E\{\tilde{I}_i(\beta_{\alpha 0})\} = \alpha$ as $n \to \infty$ and the test statistic (7) is asymptotically equivalent to the test statistic (5).

■

Derivation of the nonparametric saddlepoint test statistic

To compute the saddlepoint test statistic, we follow the result given for $M-$estimators in the Appendix. Here the n distributions F^i are estimated through the empirical distribution of the residuals as in Ronchetti & Welsh (1994).

Let us define

$$\begin{aligned}
r_i &= y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_\alpha \\
I_{ij}(\boldsymbol{\beta}) &= \mathrm{I}[r_i + \mathbf{x}_j^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}) < 0] = \mathrm{I}[y_i - \mathbf{x}_j^\top \boldsymbol{\beta} < 0] \\
F_n^j(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^n I_{ij}(\boldsymbol{\beta}), \qquad i,j = 1,\dots,n.
\end{aligned}$$

As in the parametric case,

$$\mathbf{x}_j^\top \boldsymbol{\mu} = \log \frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}$$

solves

$$\sum_{i=1}^n\sum_{j=1}^n (\alpha - \mathrm{I}[r_i + \mathbf{x}_j^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0})])\mathbf{x}_j e^{(\alpha - \mathrm{I}[r_i + \mathbf{x}_j^\top(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0})])\mathbf{x}_j^\top \boldsymbol{\mu}} = 0,$$

because the following equalities are equivalent.

$$\sum_{i=1}^n\sum_{j=1}^n (\alpha - I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha - I_{ij}(\boldsymbol{\beta}_{\alpha 0})}\mathbf{x}_j = 0$$

$$\iff \sum_{i=1}^n\sum_{j=1}^n \alpha\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha - I_{ij}(\boldsymbol{\beta}_{\alpha 0})}\mathbf{x}_j$$

$$= \sum_{i=1}^n\sum_{j=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha 0})\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha - I_{ij}(\boldsymbol{\beta}_{\alpha 0})}\mathbf{x}_j$$

$$\iff \sum_{i=1}^n\sum_{j=1}^n \alpha I_{ij}(\boldsymbol{\beta}_{\alpha 0})\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{1-\alpha}\mathbf{x}_j$$

$$+ \sum_{i=1}^n\sum_{j=1}^n \alpha(1 - I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j$$

$$= \sum_{i=1}^n\sum_{j=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha 0})\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{1-\alpha}\mathbf{x}_j$$

$$\iff \sum_{i=1}^n\sum_{j=1}^n \alpha(1 - I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(1-\alpha)I_{ij}(\boldsymbol{\beta}_{\alpha 0})\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{1-\alpha}\mathbf{x}_j$$

$$\Longleftrightarrow\ n\sum_{j=1}^{n}\left(\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j - \sum_{j=1}^{n}\sum_{i=1}^{n}I_{ij}(\boldsymbol{\beta}_{\alpha 0})\left(\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{n}\frac{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\left(\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j - \sum_{j=1}^{n}\sum_{i=1}^{n}I_{ij}(\boldsymbol{\beta}_{\alpha 0})\left(\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j$$

$$\Longleftrightarrow\ n\sum_{j=1}^{n}\left(\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j = \sum_{j=1}^{n}\frac{1}{F_n^j(\boldsymbol{\beta}_{\alpha 0})}\left(\frac{F_n^j(\boldsymbol{\beta}_{\alpha 0})}{1-F_n^j(\boldsymbol{\beta}_{\alpha 0})}\right)^{\alpha}\mathbf{x}_j\sum_{i=1}^{n}I_{ij}(\boldsymbol{\beta}_{\alpha 0}),$$

which is an identity because by definition

$$\sum_{i=1}^{n}I_{ij}(\boldsymbol{\beta}_{\alpha 0}) = nF_n^j(\boldsymbol{\beta}_{\alpha 0}).$$

Then we can write

$$\begin{aligned} K_\psi(\boldsymbol{\lambda};\boldsymbol{\beta}) &= \frac{1}{n}\sum_{j=1}^{n}K_\psi^j(\boldsymbol{\lambda};\boldsymbol{\beta}) \\ &= \frac{1}{n}\sum_{j=1}^{n}\left\{\log\sum_{i=1}^{n}\left[e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^\top\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\mathbf{x}_j^\top\boldsymbol{\mu}}\right]\right. \\ &\qquad\left. -\log\sum_{i=1}^{n}e^{(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\mathbf{x}_j^\top\boldsymbol{\mu}}\right\}, \\ \frac{\partial K^j(\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial\boldsymbol{\lambda}} &= \frac{\partial}{\partial\boldsymbol{\lambda}}\left\{\log\sum_{i=1}^{n}e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^\top\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\mathbf{x}_j^\top\boldsymbol{\mu}}\right\} \\ &= \frac{\sum_{i=1}^{n}(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^\top\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\mathbf{x}_j^\top\boldsymbol{\mu}}}{\sum_{i=1}^{n}e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^\top\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\mathbf{x}_j^\top\boldsymbol{\mu}}}. \end{aligned}$$

Therefore, the test statistic is given by

$$2n\hat{h}(\hat{\boldsymbol{\beta}}_\alpha) = -2nK_\psi(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha);\hat{\boldsymbol{\beta}}_\alpha),$$

where $\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha)$ satisfies $\frac{\partial K_\psi^j(\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial\boldsymbol{\lambda}} = 0$, i.e.

$$\sum_{i=1}^{n}(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^\top\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha 0}))\mathbf{x}_j^\top\boldsymbol{\mu}} = 0.$$

Now let

$$I_i = \mathrm{I}[r_i < 0], \qquad i = 1, \ldots, n.$$

Then $\mathbf{x}_j^\top \boldsymbol{\lambda}$ is a solution of the equation

$$\sum_{i=1}^{n} (\alpha - I_i) e^{(\alpha - I_i)\mathbf{x}_j^\top \boldsymbol{\lambda}} \left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})} = 0$$

i.e.

$$(1-\alpha) \sum_{i=1}^{n} I_i e^{(\alpha-1)\mathbf{x}_j^\top \boldsymbol{\lambda}} \left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}$$

$$= \alpha \sum_{i=1}^{n} (1 - I_i) e^{\alpha \mathbf{x}_j^\top \boldsymbol{\lambda}} \left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}$$

and this is equivalent to

$$\mathbf{x}_j^\top \boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha) = \log \left\{ \frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij}(1 - I_i)} \right\},$$

where

$$w_{ij} = \frac{\left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}}{\sum_{k=1}^{n} \left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha 0})}{F_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{kj}(\boldsymbol{\beta}_{\alpha 0})}}.$$

Thus, we obtain

$$\begin{aligned}
K_\psi^j(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha); \hat{\boldsymbol{\beta}}_\alpha) &= \log \sum_{l=1}^{n} \left(\frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij}(1 - I_i)} \right)^{\alpha - I_l} w_{lj} \\
&= \log \Bigg\{ \left(\frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij}(1 - I_i)} \right)^{\alpha} \\
&\quad \times \sum_{l=1}^{n} \left(\frac{\alpha}{1-\alpha} \frac{\sum_{i=1}^{n} w_{ij}(1 - I_i)}{\sum_{i=1}^{n} w_{ij} I_i} w_{lj} I_l + w_{lj}(1 - I_l) \right) \Bigg\} \\
&= \log \left\{ \left(\frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^{n} w_{ij} I_i}{\sum_{i=1}^{n} w_{ij}(1 - I_i)} \right)^{\alpha} \frac{\sum_{i=1}^{n} w_{ij}(1 - I_i)}{1-\alpha} \right\} \\
&= \log \left\{ \left(\frac{\sum_{i=1}^{n} w_{ij} I_i}{\alpha} \right)^{\alpha} \left(\frac{\sum_{i=1}^{n} w_{ij}(1 - I_i)}{1-\alpha} \right)^{1-\alpha} \right\}
\end{aligned}$$

and

$$\hat{h}(\hat{\boldsymbol{\beta}}_\alpha) = -K_\psi(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha); \hat{\boldsymbol{\beta}}_\alpha) = -\frac{1}{n}\sum_{j=1}^{n} K_\psi^j(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha); \hat{\boldsymbol{\beta}}_\alpha) \tag{13}$$

$$= -\frac{1}{n}\sum_{j=1}^{n} \log\left\{\left(\frac{\sum_{i=1}^{n} w_{ij} I_i}{\alpha}\right)^{\alpha}\left(\frac{\sum_{i=1}^{n} w_{ij}(1-I_i)}{1-\alpha}\right)^{1-\alpha}\right\} \tag{14}$$

and this concludes the computation of the saddlepoint test statistic.

The nonparametric test statistic for the special case of simple quantiles can be easily obtained from (14) by putting $p = 1, x_i \equiv 1, r_i = y_i - \hat{\beta}_\alpha, \hat{\beta}_\alpha = F_n^{-1}(\alpha)$, the empirical quantile, $I_i = \mathrm{I}[r_i < 0] = \mathrm{I}[y_i < \hat{\beta}_\alpha], I_{ij}(\beta) = \mathrm{I}[y_i < \beta] \equiv \tilde{I}_i(\beta)$, and by observing that in this case the weights $w_{ij} \equiv w_i$ are independent of j.

Received 2014
Accepted